Difference Between Means

One of the most common inference problems is the following:

- We have a random sample of size m, $X_1, X_2, \ldots, X_m$ from population 1
- We have a random sample of size n, $Y_1, Y_2, \ldots, Y_n$ from population 2
- Population 1 has mean μ_x and variance σ_x^2
- Population 2 has mean μ_y and variance σ_y^2
- We are interested in comparing μ_x and μ_y

Difference Between Means

A typical problem of this type would be:

Tomato garden 1 is fertilized with plant food 1, and tomato garden 2 with plant food 2. The 25 plants in garden 1 produce an average of 3.4 pounds with a standard deviation of 1.3 pounds, while the 22 plants in garden 2 produce an average of 3.1 pounds with a standard deviation of 1.7 pounds.

Can we conclude based on these samples that tomato plants fed with plant food 1 produce more than those fed by plant food 2?

Difference Between Means

A typical problem of this type would be:

Tomato garden 1 is fertilized with plant food 1, and tomato garden 2 with plant food 2. The 25 plants in garden 1 produce an average of 3.4 pounds with a standard deviation of 1.3 pounds, while the 22 plants in garden 2 produce an average of 3.1 pounds with a standard deviation of 1.7 pounds.

Can we conclude based on these samples that tomato plants fed with plant food 1 produce more than those fed by plant food 2?

One might be tempted to jump to the conclusion that because the average yield per plant was higher for plant food 1, it is superior to plant food 2.

Difference Between Means

This simplistic approach is risky because if the two plant foods are equally effective, that is, the average difference is zero, the measured difference in an experiment will usually not be exactly zero.

Difference Between Means

This simplistic approach is risky because if the two plant foods are equally effective, that is, the average difference is zero, the measured difference in an experiment will usually not be exactly zero.

If the two plant foods are equally effective and the sample sizes reasonably large,

- About half the time plant food 1 will appear to be better
- About half the time plant food 2 will appear to be better

Difference Between Means

This simplistic approach is risky because if the two plant foods are equally effective, that is, the average difference is zero, the measured difference in an experiment will usually not be exactly zero.

If the two plant foods are equally effective and the sample sizes reasonably large,

- About half the time plant food 1 will appear to be better
- About half the time plant food 2 will appear to be better

The important question is, does a difference of $3.4 - 3.1 = 0.3$ pounds per tomato plant represent a significant difference? Or could it be attributed to sampling error?

Difference Between Means

To approach the question systematically, we will treat the quantity

$$\overline{x} - \overline{y}$$

as a random variable, representing a single observation from some probability distribution.

Difference Between Means

To approach the question systematically, we will treat the quantity

$$\overline{x} - \overline{y}$$

as a random variable, representing a single observation from some probability distribution.

Because $\overline{x}$ and $\overline{y}$ are sample means, we can state that:

$$E(\overline{x} - \overline{y}) = \mu_x - \mu_y$$

and

$$\sigma_{\overline{x}-\overline{y}} = \sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}$$

based on the properties of means and standard deviations.

Difference Between Means

If we assume that the underlying populations are normal, the problem of comparing the unknown means when we do not know the population variances is called the *Fisher-Behrens* problem.

Difference Between Means

If we assume that the underlying populations are normal, the problem of comparing the unknown means when we do not know the population variances is called the *Fisher-Behrens* problem.

It is a rather suprising fact that *there is no completely satisfactory solution* to this problem.

Difference Between Means

If we assume that the underlying populations are normal, the problem of comparing the unknown means when we do not know the population variances is called the *Fisher-Behrens* problem.

It is a rather suprising fact that *there is no completely satisfactory solution* to this problem.

With that said, the techniques we will cover are widely used and generally considered the best way to handle this problem.

Difference Between Means

We will consider the problem of comparing the means of two populations in four cases:

- The case in which σ_x and σ_y are known
- The case in which σ_x and σ_y are unknown
- The case in which we are comparing proportions
- The case in which the measurements are paired for each subject

Difference Between Means

We will consider the problem of comparing the means of two populations in four cases:

- The case in which σ_x and σ_y are known
- The case in which σ_x and σ_y are unknown
- The case in which we are comparing proportions
- The case in which the measurements are paired for each subject

The last case is different from the first three in that we can reduce it to the single sample problem by considering the difference $X_i - Y_i$ for each subject. This model usually arises from taking measurements at two different times on each subject in the sample.

Known Sigma

In this case, the difference between the sample means $\overline{x} - \overline{y}$ has a normal distribution, so once again we are working with bell curves.

Known Sigma

In this case, the difference between the sample means $\overline{x} - \overline{y}$ has a normal distribution, so once again we are working with bell curves.

When the population variances σ_x^2 and σ_y^2 are known,

$$\overline{x} - \overline{y} \sim N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}\right)$$

Known Sigma

In this case, the difference between the sample means $\overline{x} - \overline{y}$ has a normal distribution, so once again we are working with bell curves.

When the population variances σ_x^2 and σ_y^2 are known,

$$\overline{x} - \overline{y} \sim N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}\right)$$

We will use the author's guidline that as long as both samples are larger than 40, we can use this form even when the sample standard deviations s_x and s_y are used in place of σ_x and σ_y.

Known Sigma

The SAT test is administered to 87 students at school 1 and 76 students at school 2. The mean score at school 1 was 498.3, while the mean score at school 2 was 504.3. Test the hypothesis that the mean SAT score at the two schools is the same against the alternative that the two schools are different at the level $\alpha = 0.05$.

Known Sigma

The SAT test is administered to 87 students at school 1 and 76 students at school 2. The mean score at school 1 was 498.3, while the mean score at school 2 was 504.3. Test the hypothesis that the mean SAT score at the two schools is the same against the alternative that the two schools are different at the level $\alpha = 0.05$.

Use the σ Known link in the *Inference About Two Means* section of the technology page because the SAT is standardized to have $\sigma = 100$.

Known Sigma

The SAT test is administered to 87 students at school 1 and 76 students at school 2. The mean score at school 1 was 498.3, while the mean score at school 2 was 504.3. Test the hypothesis that the mean SAT score at the two schools is the same against the alternative that the two schools are different at the level $\alpha = 0.05$.

Use the σ Known link in the *Inference About Two Means* section of the technology page because the SAT is standardized to have $\sigma = 100$.

Fill in α=0.05 and:

- $xbar1$=498.3, σ_1=100, n_1=87
- $xbar2$=504.3, σ_2=100, n_2=76

Known Sigma

A researcher wishes to determine whether the presence of entomophthoralean fungus (Entomophaga maimaiga), indicates a reduced level of gypsy moth (Lymantria dispar) infestation. Gypsy moth egg masses are counted in 43 quarter-acre plots where the fungus is detected, and 52 where it is not. The number of egg masses in plots where the fungus is detected has a sample mean of 32.4 and a sample standard deviation of 12.1. In plots where no fungus is detected, the sample mean is 43.1 with a sample standard deviation of 15.0. Does this study support the claim that the fungus lowers the level of gypsy moth infestation?

Known Sigma

A researcher wishes to determine whether the presence of entomophthoralean fungus (Entomophaga maimaiga), indicates a reduced level of gypsy moth (Lymantria dispar) infestation. Gypsy moth egg masses are counted in 43 quarter-acre plots where the fungus is detected, and 52 where it is not. The number of egg masses in plots where the fungus is detected has a sample mean of 32.4 and a sample standard deviation of 12.1. In plots where no fungus is detected, the sample mean is 43.1 with a sample standard deviation of 15.0. Does this study support the claim that the fungus lowers the level of gypsy moth infestation?

Use the σ Known link in the *Inference About Two Means* section of the technology page because both samples are larger than 40.

Known Sigma

Fill in α=0.05 and:

- $xbar1$=32.4, σ_1=12.1, n_1=43
- $xbar2$=43.1, σ_2=15.0, n_2=52

Unknown Sigma

In this case, we assume we are sampling from a normal population, but we do not know the population standard diviations and at least one sample is 40 or smaller.

Unknown Sigma

In this case, we assume we are sampling from a normal population, but we do not know the population standard diviations and at least one sample is 40 or smaller.

The test statistic in this case is *approximately* distributed as a t statistic.

Unknown Sigma

In this case, we assume we are sampling from a normal population, but we do not know the population standard diviations and at least one sample is 40 or smaller.

The test statistic in this case is *approximately* distributed as a t statistic.

The appropriate degrees of freedom for the t distribution is computed using a complicated formula. When computed in the manner outlined in the text, the statistic is known as *Welch's approximate t*.

Unknown Sigma

In an experiment to evaluate the effect of high temperature and humidity on stored films, 35 identical test films are stored in a climate-controlled environment while another 50 identical test films are exposed to ambient temperature and humidity. After one year, the films are evaluated for loss of transparency using a scale that runs from 0 to 100, 100 indicating no measurable loss of transparency and 0 indicating complete loss of transparency. The average measure for the climate-controlled films is 98.2 with a standard deviation of 0.3. The average measure for the non climate-controlled films is 97.3 with a standard deviation of 0.5. Does this data indicate that climate control is effective in preserving transparency?

Unknown Sigma

In an experiment to evaluate the effect of high temperature and humidity on stored films, 35 identical test films are stored in a climate-controlled environment while another 50 identical test films are exposed to ambient temperature and humidity. After one year, the films are evaluated for loss of transparency using a scale that runs from 0 to 100, 100 indicating no measurable loss of transparency and 0 indicating complete loss of transparency. The average measure for the climate-controlled films is 98.2 with a standard deviation of 0.3. The average measure for the non climate-controlled films is 97.3 with a standard deviation of 0.5. Does this data indicate that climate control is effective in preserving transparency?

Use the σ Unknown link in the *Inference About Two Means* section of the technology page because one sample is smaller than 41.

Unknown Sigma

Fill in α=0.05 and:

- $xbar1$=98.2, s_1=0.3, n_1=35
- $xbar2$=97.3, s_2=0.5, n_2=50

Unknown Sigma

A government agency is trying to build a case to support the assertion that ocean dumping of titanium dioxide waste in the Baltimore Canyon is resulting in detectable increases in titanium in the deep sea scallop (Placopecten magellanicus) population in the dumping area. A sample of 45 specimens from the dumping area showed titanimum levels of 12.3 parts per million with a standard deviation of 5.1, while a sample of 38 specimens from a site 20 miles away had an average concentration of 8.6 parts per million with a standard deviation of 4.2. Determine whether this data supports the agency's assertion.

Unknown Sigma

A government agency is trying to build a case to support the assertion that ocean dumping of titanium dioxide waste in the Baltimore Canyon is resulting in detectable increases in titanium in the deep sea scallop (Placopecten magellanicus) population in the dumping area. A sample of 45 specimens from the dumping area showed titanimum levels of 12.3 parts per million with a standard deviation of 5.1, while a sample of 38 specimens from a site 20 miles away had an average concentration of 8.6 parts per million with a standard deviation of 4.2. Determine whether this data supports the agency's assertion.

Use the σ Unknown link in the *Inference About Two Means* section of the technology page because one sample is smaller than 41.

Unknown Sigma

Fill in α=0.05 and:

- $xbar1$=12.3, s_1=5.1, n_1=45
- $xbar2$=8.6, s_2=4.2, n_2=38

Proportions

In this case, we assume our two sample means represent the *proportions* of the samples that have a certain characteristic.

Proportions

In this case, we assume our two sample means represent the *proportions* of the samples that have a certain characteristic.

The number of individual in each sample X and Y that have the characteristic can be thought of as binomial random variables. We use a large sample normal approximation to compute the test statistic.

Proportions

In this case, we assume our two sample means represent the *proportions* of the samples that have a certain characteristic.

The number of individual in each sample X and Y that have the characteristic can be thought of as binomial random variables. We use a large sample normal approximation to compute the test statistic.

In the case where the null hypothesis is that the proportions are equal, the test statistic is

$$z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{m}+\frac{1}{n}\right)}} \qquad \hat{p} = \frac{X+Y}{m+n}$$

Proportions

A sample of 85 deer ticks (Ixodes scapularis) from Long Island contains 12 that test positive for the Ehrlichia chaffeensis bacteria. A sample of 120 deer ticks collected on Cape Cod contains 14 that test positive. Does the sample evidence support the claim that the level of Ehrlichia chaffeensis infection at the collection sites is identical?

Proportions

A sample of 85 deer ticks (Ixodes scapularis) from Long Island contains 12 that test positive for the Ehrlichia chaffeensis bacteria. A sample of 120 deer ticks collected on Cape Cod contains 14 that test positive. Does the sample evidence support the claim that the level of Ehrlichia chaffeensis infection at the collection sites is identical?

Use the Proportions link in the *Inference About Two Means* section of the technology page.

Proportions

Fill in α=0.05 and:

- $p\ hat1$=12/85, n_1=85
- $p\ hat2$=14/120, n_2=120

Proportions

Fill in α=0.05 and:

- $p\ hat1$=12/85, n_1=85
- $p\ hat2$=14/120, n_2=120

Note the spreadsheet includes a check for one criterion that the sample size is large enough to use the normal approximation:

$$n \cdot (\hat{p}(1 - \hat{p}) > 10$$

as well as an $= IF$ statement to display "Valid" or "Not Valid" using this rule.

Proportions

In a double-blind study of Major Depressive Disorder (MDD), 75 subjects are treated with seratonin-reuptake inhibitors (SRIs) while 85 are given a placebo. After 8 weeks of treatment, 18 subjects in the SRI group have experienced a remission of MDD, while 12 subjects in the placebo group have. Can we conclude that patients receiving the drug are more likely to remit than those receiving a placebo?

Proportions

In a double-blind study of Major Depressive Disorder (MDD), 75 subjects are treated with seratonin-reuptake inhibitors (SRIs) while 85 are given a placebo. After 8 weeks of treatment, 18 subjects in the SRI group have experienced a remission of MDD, while 12 subjects in the placebo group have. Can we conclude that patients receiving the drug are more likely to remit than those receiving a placebo?

Use the Proportions link in the *Inference About Two Means* section of the technology page.

Proportions

Fill in α=0.05 and:

- $p\ hat1$=18/75, n_1=75
- $p\ hat2$=12/85, n_2=85

Paired or Dependent Samples

In this case, we assume our sample consists of repeated measurements on the subjects. This occurs in a "before and after" scenario, where subjects are measured before and after some treatment occurs.

It also occurs in various economic measures such as "same store sales".

Paired or Dependent Samples

In this case, we assume our sample consists of repeated measurements on the subjects. This occurs in a "before and after" scenario, where subjects are measured before and after some treatment occurs.

It also occurs in various economic measures such as "same store sales".

The paired data or repeated measures design is usually better at detecting differences than an unpaired design because the pairing tends to eliminate the effects of uncontrolled variables.

Paired or Dependent Samples

In this case, we effectively reduce the problem to a single sample by combining pairs of values X_i, Y_i for each subject into a single difference $d_i = X_i - Y_i$ and then treating the d_i values as the sample.

The test statistic is:

$$t = \frac{\overline{d} - \Delta_0}{s_d/\sqrt{n}}$$

Paired or Dependent Samples

Mosquito traps are placed near 43 small ponds and a count of Culex species in the traps is obtained during a baseline period. At the end of the baseline period a spraying program is conducted. One week after the spraying, the traps are cleaned and a second collection period is initiated. Based on an estimate of the size of each pond, the raw counts are converted to a density of Culex mosquitoes per square foot of pond. The difference between the before and after densities is found to have a sample mean of -4.1 and a sample standard deviation of 12.0. Test whether or not the data indicates that the spraying was effective in reducing the density of Culex species mosquitoes.

Paired or Dependent Samples

Mosquito traps are placed near 43 small ponds and a count of Culex species in the traps is obtained during a baseline period. At the end of the baseline period a spraying program is conducted. One week after the spraying, the traps are cleaned and a second collection period is initiated. Based on an estimate of the size of each pond, the raw counts are converted to a density of Culex mosquitoes per square foot of pond. The difference between the before and after densities is found to have a sample mean of -4.1 and a sample standard deviation of 12.0. Test whether or not the data indicates that the spraying was effective in reducing the density of Culex species mosquitoes.

Use the Paired or Dependent Samples link in the *Inference About Two Means* section of the technology page.

Paired or Dependent Samples

Fill in α=0.05 and:

- d_bar=-4.1, $s_d =$=12.0 n_1=43

Paired or Dependent Samples

In a study of the effect of long-term talk therapy on depression, 48 subjects diagnosed with major depressive disorder (MDD) are given a modified version of the Beck Depression Inventory (BDI) after six months and again after nine months of treatment. The mean difference between the six month and nine month BDI scores was -1.0 with a standard deviation of 3.9. Does this data support the claim that talk therapy beyond six months is not effective.

Paired or Dependent Samples

In a study of the effect of long-term talk therapy on depression, 48 subjects diagnosed with major depressive disorder (MDD) are given a modified version of the Beck Depression Inventory (BDI) after six months and again after nine months of treatment. The mean difference between the six month and nine month BDI scores was -1.0 with a standard deviation of 3.9. Does this data support the claim that talk therapy beyond six months is not effective.

Use the Paired or Dependent Samples link in the *Inference About Two Means* section of the technology page.

Paired or Dependent Samples

Fill in α=0.05 and:

- d_bar=-1.0, s_d=3.9 n_1=48